Minutes of the Database and Related Utilities meeting for 12/20/02:

Attending: Joanne Bogart, Pat Nolan, Bob Schaefer

Minutes:

Bob: Karl Young and Dave Davis are out, Traudl's PC is dead so she won't be joining us, so this will be a short meeting.

Bob: I have been out with walking pneumonia, so I have gotten behind in my action items. I will summarize what I have been up to,

1) The meeting time may move to Wednesday at the same time.
2) The GSFC Beowulf cluster that I used for the prototype D1 has fallen into a limbo state. It is owned by the Lab for High Energy Astrophysics, but is physically located in another building. Our systems administrators don't want anything to do with it unless it is located here. Lab interest has waned in using it (it is after all a set of 400Mhz Intel Celeron processors with 17Gb disks). The group who is housing the cluster do not want to pay for any system admin support, so no one with root access wants to touch it to modify or add accounts to it. Since it is a political problem, perhaps it is best just to try to use the SLAC SCS machines to prototype.
3) I have sent around a strawman design for D1. It is now linked to the agenda page and perhaps we can discuss when more people are here.
4) Traudl has sent us a requirements document for the mirrorsite. It is linked to the agenda page as well.

Anyone see anything in Traudl's document that they wanted to comment on?

Pat: I took a quick look at Traudl's document and nothing of concern seemed to jump out at me.

Bob: There were a few questions contained in the document that should be answered by us, so we should probably look at it carefully. Also there is an old diagram taken from my June Workshop Beowulf presentation which had the photons kept in memory. (Note added after the meeting by Bob - I am not sure we want to do this. The SCS people thought this would make our system too complicated, and I tend to agree.) I attempted to get my PERL D1 prototype to load the photons in memory. The data took 2-4 times more memory space than the files did on disk. (Note added after meeting by Bob – it was also 3-4 times SLOWER than the version that read files from disk), At this point I gave up and decided the whole thing should be written in C, partly because the Perl wrapper around cfitsio misses some very useful functions and the wrapper around MPI is flaky.

Pat: Perl and scripting languages don't do well with arrays, because they are not set up to deal with them. Since they are not typed, perl carries around some overhead with variables so it can figure out what to do with them on the fly. A perl array therefore is not going to be nearly as efficient with memory as a typed stored array in a compiled

language. There has been some work with a scientific version of perl that tries to define numeric type arrays.

Bob: Perl does support binary types. I tried to read in the table rows as binary and it still used > 2 times more memory than the file size. Plus parsing them is a pain because the pack and unpack use the native little-endian while FITS files are in network byte order (big-endian), so that was another reason I wanted to give up on perl. However that does sound interesting. Do you know if Python has the same problem?

Pat: I believe it does.

Bob: I think this is worth following up because it would seem useful to know this for the User Interface group.

Bob: Well, given that we are missing so many people from the meeting, I see no reason to continue. I will be in contact to set up the first meeting next year. In the meantime, have a good holiday and see you in the new year!

Action Items:

This week:
Bob – check out array handling in Perl for scientific programming.

Holdovers:

Bob

Continue to pursue finding hardware for testing (both the LHEA cluster and our own machines).
Post web page about MPI vs. PVM.
Collect responses about requested queries and catalogs.
Continue investigating HEASARC for allowed methods of serving databases from the web
Contact Richard about moving DB web pages to SLAC (get citrix working.)

Karl:

Post a web page about cluster management software.
Continue efforts to get hardware for testing.
Circulate strawman database designs.

Dave D.

Continue efforts to get hardware for testing.

Toby Burnett

Send summary of work done on D2 to e-mail distribution list.